Computational Economics (2025) 65:2227–2248
https://doi.org/10.1007/s10614-024-10588-3

Bitcoin Price Prediction Using Sentiment Analysis and Empirical Mode Decomposition

Serdar Arslan[1]

Accepted: 6 March 2024 / Published online: 28 May 2024

Abstract

Cryptocurrencies have garnered significant attention recently due to widespread investments. Additionally, researchers have increasingly turned to social media, particularly in the context of financial markets, to harness its predictive capabilities. Investors rely on platforms like Twitter to analyze investments and detect trends, which can directly impact the future price movements of Bitcoin. Understanding and analyzing Twitter sentiments can potentially provide insights into future Bitcoin price movements and can shed light on how investor sentiment affects cryptocurrency markets. In this study, we explore the correlation between Twitter activity and Bitcoin prices by examining tweets related to Bitcoin price sentiments. Our proposed model consists of two distinct networks. The first network exclusively utilizes historical price data, which is further decomposed into various components using the Empirical Mode Decomposition method. This decomposition helps mitigate the impact of irregular fluctuations on Bitcoin price predictions. Each of these components is then separately processed by Long Short-Term Memory (LSTM) networks. The second network focuses on modeling user sentiments and emotions in conjunction with Bitcoin market data. User opinions are categorized into positive and negative classes and are integrated with historical data to predict the next-day price using LSTM networks. Finally, the outputs of each network are combined to form the ultimate prediction values. Experimental results demonstrate that Twitter sentiment can effectively helps us predict Bitcoin price trends. Furthermore, to validate our proposed model, we compared it with several state-of-the-art methods. The results indicate that our approach outperforms these existing models in terms of accuracy.

Keywords Cryptocurrency · Prediction · EMD · Ensemble learning

✉ Serdar Arslan
sarslan@cankaya.edu.tr

[1] Computer Engineering Department, Cankaya University, 06790 Etimesgut, Ankara, Turkey

1 Introduction

The surge in the popularity of cryptocurrencies has created a growing demand for accurate price predictions. Investors in these digital currencies are constantly seeking insights into price trends to make well-informed investment decisions. However, forecasting cryptocurrency prices is a formidable task, primarily due to their extreme volatility and susceptibility to external factors like news and social media. In recent years, researchers have devised a range of methods for predicting cryptocurrency prices, spanning from conventional statistical approaches to advanced machine learning algorithms. One of the key challenges in this endeavor is the availability of reliable data. Nonetheless, the emergence of web scraping techniques has made it increasingly convenient to gather data from diverse sources, including news websites and social media platforms (Kraaijeveld & De Smedt, 2020).

Cryptocurrencies have experienced rapid growth and widespread adoption, with increasing utilization in official financial transactions and the exchange of goods. This surge in cryptocurrency activity, coupled with the availability of high-frequency data, has prompted the adoption of deep learning (DL) techniques, particularly in the realm of price prediction. Price prediction is a critical element of financial decision-making, encompassing portfolio optimization, risk assessment, and trading strategies (Ji et al., 2019).

The inherent heterogeneity of cryptocurrency datasets presents a challenge for traditional machine learning models. Furthermore, the substantial price fluctuations and volatility characteristic of cryptocurrencies significantly impact trading strategies and investment choices (Gurrib & Kamalov, 2022). Hence, there is a pressing need to develop models that can accurately forecast cryptocurrency prices, leveraging real-time price movement information to yield substantial profits and minimize investment risks for stakeholders.

The widespread popularity of cryptocurrencies, with Bitcoin leading the market in terms of volume, has drawn the attention of numerous investors. In 2021, Bitcoin reached multiple unprecedented all-time high prices, leading to a surge in the number of individuals speculating on Bitcoin's price movements via social media platforms. Twitter, in particular, serves as a valuable source of instant information and updates on cryptocurrencies. Moreover, Twitter offers a wealth of emotional intelligence, as investors often express their sentiments and emotions related to cryptocurrency investments on the platform. It is worth noting that these sentiments and emotions can significantly influence investment strategies and the decision-making process, as highlighted in a study by Kraaijeveld and De Smedt (2020).

The popularity of Bitcoin may cause several drawbacks and large price fluctuations with high volatility can be seen as major problem (Khedr et al., 2021). Moreover, the Bitcoin market volatility mainly emerged from news messages and social media posts (Kraaijeveld & De Smedt, 2020). Thus, it should be noted that forecasting cryptocurrencies' prices are fundamentally different from forecasting other financial assets because of these drawbacks (Mohapatra et al., 2019).

Current global regulations pertaining to cryptocurrencies are notably sparse, primarily because cryptocurrencies have yet to be officially recognized as a fully developed asset class. This regulatory gap, coupled with the immense popularity of cryptocurrencies and the absence of an institutional authority, has resulted in the cryptocurrency market's extreme volatility. Therefore, cryptocurrency prices may not be treated as traditional currencies in this context, and the prediction of the price of any cryptocurrency is still a challenging problem (Gurrib & Kamalov, 2022).

Bitcoin price data can actually be seen as a clear example of time-series data. Thus, prediction of this price is a simple time series forecasting problem (Roy et al., 2018). Time series forecasting presents a significant challenge, involving the utilization of past values of a dependent variable to predict its future values. Within the academic literature, time series forecasting finds applications across diverse domains, including energy, business, economics, healthcare, and the environment. Notably, finance time series forecasting stands out as one of the most extensively researched areas (Shin et al., 2021).

Traditionally, statistical techniques like Moving Average (MA), Auto-Regression (AR), Auto-Regressive Integrated Moving Average (ARIMA), as well as machine learning methods such as Support Vector Machine (SVM), Decision Trees, and Artificial Neural Network (ANN), have been employed in financial forecasting. However, with the rise of machine learning and deep learning approaches, Recurrent Neural Networks (RNNs) have gained prominence in recent times for forecasting tasks, demonstrating commendable performance (Lim & Zohren, 2021).

The cryptocurrency market's high volatility is significantly influenced by news articles and social media posts. This impact is further amplified as investors grapple with discerning the authenticity of the information shared. Given the relatively young nature of the cryptocurrency market, traditional news outlets may not always report events promptly, leading cryptocurrency investors to rely heavily on social media as their primary information source (Giachanou & Crestani, 2016).

In particular, the micro-blogging platform Twitter stands out as a widely utilized source for cryptocurrency-related information. Twitter not only offers real-time updates on cryptocurrencies but also serves as a valuable repository of emotional intelligence. Investors frequently express their sentiments on this platform. Behavioral economics teaches us that sentiment and emotions can profoundly shape individual behavior and decision-making processes. Given the abundant and readily accessible data from Twitter, which includes the emotional insights of cryptocurrency users and investors, this study's primary objective is to investigate the extent to which public sentiment on Twitter can be harnessed to predict fluctuations in cryptocurrency prices (Kraaijeveld & De Smedt, 2020).

To effectively predict the price of any cryptocurrency, the literature has extensively employed various traditional time-series forecasting methods. Additionally, it is crucial to investigate the influence of user sentiments and emotions on Bitcoin price trends, which necessitates the analysis of data gathered from diverse sources like social media, blogs, forums, and more. As a result, tweets related to cryptocurrencies should undergo scrutiny through the application of sentiment analysis techniques (Cocco et al., 2021).

Sentiment analysis represents one of the most prominent applications of Natural Language Processing (NLP). It entails a straightforward classification task, where text-based opinions are categorized as positive, negative, or neutral. Twitter, in particular, serves as a valuable corpus for sentiment analysis across a multitude of applications, including news, stock market analysis, assessment of product brands, evaluation of presidential candidate performances in debates or elections, and more (Giachanou & Crestani, 2016; Zimbra et al., 2018).

Consequently, the primary motivation behind this study is to explore the feasibility of predicting the closing prices of cryptocurrencies using an optimized model with Twitter sentiment analysis using deep learning techniques and empirical mode decomposition. In this work, Bitcoin price prediction is achieved using an ensemble model combining LSTM with tweet sentiments. The model consists of two parts; the first for price and opinions and the second for only prices. For the sentiment part of the proposed model, unsupervised sentiment analysis is applied to tweets collected in the one-year interval. The positive and negative ratios of daily tweets are evaluated and combined with other features of Bitcoin, such as market volume, close price, and open price. On the other hand, for the second part of the proposed model, an empirical mode decomposition (EMD)-based deep learning approach which combines the EMD method with the long short-term memory network model to estimate Bitcoin price is modeled. For this purpose, the EMD algorithm decomposes Bitcoin daily prices into several intrinsic mode functions (IMFs). Then, an LSTM model is trained separately for each extracted IMFs. Finally, the prediction results of all IMFs are combined by summation to determine an aggregated output for Bitcoin price prediction. Finally, the outcomes of these two parts are aggregated, and final prediction results are obtained. Thus the main contributions of our work can be summarized as follows;

- Investigating the challenge of forecasting Bitcoin prices for the following day and the impact of tweets to this price involved the application of a deep learning approach known as the Long Short-Term Memory (LSTM) model.
- Utilization of Empirical Mode Decomposition (EMD) as a data decomposition technique aids in the discrimination of high and low-frequency components, enabling a comprehensive examination of the attributes inherent in each component.
- Generating a real-world dataset consisting of Tweets related to Bitcoin and subsequently processing it for analysis.

The rest of the paper is organized as follows: Sect. 2 gives information about related studies and a comparison of this study with some of the existing ones. In Sect. 3, the proposed model is explained. Details of the experiments and analysis results performed with the dataset are given in Sect. 4, and finally, in Sect. 5 we give our conclusions.

2 Related Work

Time-series forecasting and modeling have been popular research area for decades. Traditional methods have been used predominantly for linear time series for different domains. Recently, with an increase in data amount and computing power,

deep learning techniques such as LSTMs and GRUs have gained popularity in this research area (Fischer & Krauss, 2018; Keceli et al., 2020; Lara-Benítez et al., 2021; Lim & Zohren, 2021).

With the increasing trend and market volume of cryptocurrencies, the researchers paid attention to cryptocurrency price prediction. Munim et al. (2019) analyzed Bitcoin price forecasts using ARIMA and neural network autoregression (NNAR) models. Chen et al. (2021) used a two-stage model for Bitcoin price prediction. Random Forest (RF) model and artificial neural network (ANN) are used for feature extraction, and LSTM is modeled with these features for final predictions. Lahmiri et al. (2020) applied a combination of two models; radial basis function neural networks (RBFNN) and generalized regression neural networks (GRNN). An SVM-based price prediction model is discussed in Zhao et al. (2019). Gyamerah (2019) analyzed various machine learning techniques such as generalized linear model via penalized maximum likelihood, random forest, support vector regression with linear kernel, and stacking ensemble model for Bitcoin price prediction using historical data. The study of Hamayel and Owda (2021) proposes GRU, LSTM, and Bidirectional LSTM models to predict the prices of three cryptocurrencies.

Traditional and machine learning methods have some limitations for cryptocurrency data since they have natural stochasticity and irregular fluctuations. Derbentsev et al. (2021), Pintelas et al. (2020). Therefore, ensemble learning methods are applied to generate efficient prediction models by defining multiple learners to deal with these irregular patterns. The main goal of these ensemble learning models is to reduce the bias and/or variance error in the prediction task (Livieris et al., 2020; Mtiraoui et al., 2023).

Empirical mode decomposition (EMD) is an unsupervised decomposition method of signals into a set of Intrinsic Mode Functions (IMFs) along with a residue (Huang et al., 1998). These IMFs represent a different feature of the original signal at different time scales. Chen et al. (2019) proposed a hybrid EMD-LSTM-based model for financial time series forecasting. Bedi and Toshniwal (2018) uses EMD with LSTM for the electricity demand prediction of India. Aggarwal et al. (2020) uses complete empirical ensemble mode decomposition (CEEMD) with SVM to analyze the nature of the Bitcoin price series. The study of Dokur et al. (2016) proposes a hybrid model based on EMD and Artificial Neural Networks (ANN) for renewable energy systems. In Jin et al. (2021), the authors try to explore how Bitcoin prices respond to uncertainties surrounding fiat currencies. To analyze this relationship, they introduce an approach based on Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) for event analysis.

Besides historical data, collecting online news or Twitter data provides new sources of predicting the trend and price of Bitcoin (Giachanou & Crestani, 2016). Bollen et al. (2011) indicates that tweets are related to the fluctuation of the stock market. The study of Kraaijeveld and De Smedt (2020) covers the predictive power of Twitter sentiment in the setting of multiple cryptocurrencies. Kristoufek (2015) studied two main factors that effects cryptocurrency prices; internal factors such as past data, mining difficulties, etc., and external factors such as trends and other economic factors. Prajapati (2020) discussed a comprehensive analysis of different simple deep learning models using social sentiments, particularly on data collected

from Google News and Reddit. In Jin et al. (2020), the authors suggested incorporating investor sentiment into stock prediction models, as it can significantly enhance prediction accuracy. They proposed a gradual decomposition of the intricate stock price sequence using EMD and LSTM. Gurrib and Kamalov (2021) developed a model that uses linear discriminant analysis (LDA) and sentiment analysis for Bitcoin price movements. LSTM-based sentiment analysis for stock price forecast is proposed in Ko and Chang (2021). They used forum entries and news articles for sentiment analysis of Chinese stock market data. In Zhang et al. (2023), the authors introduce an innovative two-stage approach based on Variational Mode Decomposition (VMD) for conducting a multi-scale regression analysis of various cryptocurrency attributes, which remain relatively unexplored in the current literature. In the initial stage, the cryptocurrency prices are decomposed using Variational Mode Decomposition (VMD) into low, medium, and high-frequency modes, each possessing distinct characteristics. In the subsequent stage, they present the VMD-based multi-scale regression, applying it to these modes along with selected explanatory variables. To demonstrate the effectiveness of this framework, they concentrate on an in-depth analysis of the multiple attributes associated with daily Bitcoin price data as a case study.

3 Methodology

This section will explain the proposed ensemble learning method. It will describe the system architecture first and then, data collection and preprocessing, the sentiment analysis, and finally ensemble model for Bitcoin price prediction in detail.

3.1 System Architecture

The overall flow diagram of the proposed model is depicted in Fig. 1. The proposed model consists of several steps. First of all, tweets are collected and cleared for

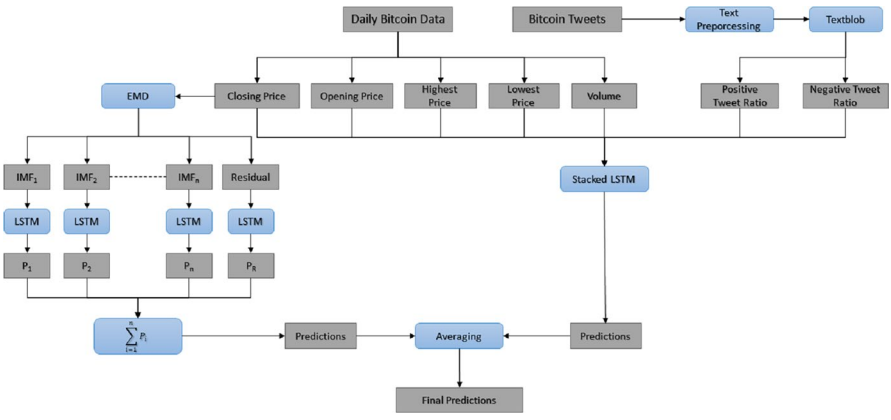

Fig. 1 The flow diagram of proposed model

classification. After applying the unsupervised classification task, the daily ratio of positive and negative tweets is calculated. In the meantime, Bitcoin's historical price and volume data are collected. The proposed model is trained along with two different models; one model is trained using decomposed Bitcoin data, and the other model uses merged input of Bitcoin data with Twitter sentiments. EMD is applied to raw Bitcoin data for the first model, and IMFs are produced along with residual. For this purpose, the work published in Zhou et al. (2022) is adapted. Each IMFs is then given as input to separate LSTM models to produce individual prediction values (EMD-LSTM). Each LSTM model's output is aggregated to make the final output of this part of the proposed model. In the second part of the model, raw Bitcoin historical data is merged with Twitter sentiments (i.e., positive and negative ratios of daily tweets), and a stacked LSTM (Stacked Sentiment LSTM - SSLSTM) is modeled using this merged input. The output of SSLSTM is averaged with those of EMD-LSTM to get final prediction results. The proposed model's steps are explained in more detail in the following sections.

3.2 Data Collection and Preprocessing

In this study, a corpus of Bitcoin related tweets are collected using the keywords "#Bitcoin", "#BTC", "#cryptocurrency" and "#crypto" (Coinlegs http://www.coinlegs.com). There are numerous hashtags in the cryptocurrency domain, but the proposed study cannot include all of them because of the computation cost. Tweets are randomly collected from users with many followers since there are some bots and fake users. Tweets are then stored in a disk for further processing. The raw data is collected between 01 January 2021 and 06 January 2022. Due to the high volume of tweets, some days are missing (25 days, approximately 0,067 percent of the real

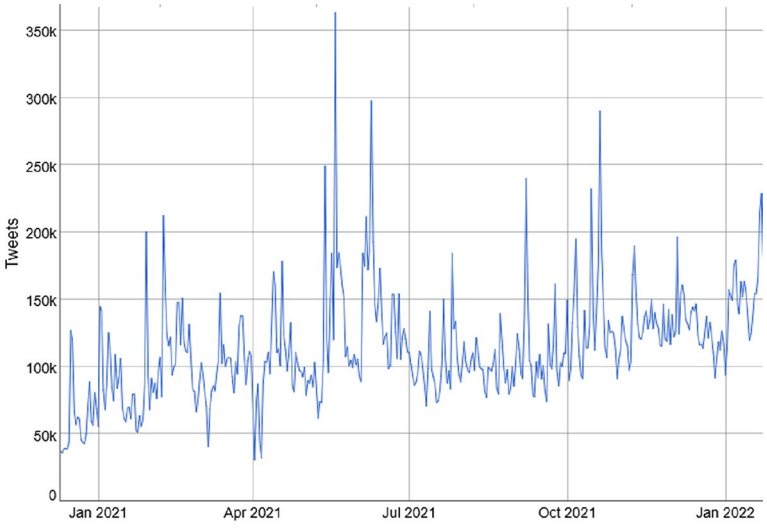

Fig. 2 Daily Bitcoin tweet count

	user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites	user_verified	date	text	hashtags	source	is_retweet
0	DeSota Wilson	Atlanta, GA	Biz Consultant, real estate, fintech, startups...	2009-04-26 20:05:09	8534.0	7605	4838	False	2021-02-10 23:59:04	Blue Ridge Bank shares halted by NYSE after #b...	['bitcoin']	Twitter Web App	False
1	CryptoND	NaN	BITCOINLIVE is a Dutch platform aimed at inf...	2019-10-17 20:12:10	6769.0	1532	25483	False	2021-02-10 23:58:48	Today, that's this #Thursday, we will do a ...	['Thursday', 'Btc', 'wallet', 'security']	Twitter for Android	False
3	Crypto is the future	NaN	I will post a lot of buying signals for BTC tr...	2019-09-28 16:48:12	625.0	129	14	False	2021-02-10 23:54:33	$BTC A big chance in a billion! Price: \487264...	['Bitcoin', 'FX', 'BTC', 'crypto']	dlvr.it	False
4	Alex Kirchmaier ATSE #FactsSuperspreader	Europa	Co-founder @RENJERJerky | Forbes 30Under30 | I...	2016-02-03 13:15:55	1249.0	1472	10482	False	2021-02-10 23:54:06	This network is secured by 9 508 nodes as of t...	['BTC']	Twitter Web App	False
5	ZenBenz™ ₿ ✪ 20732	Bkk, Thailand	I'm a cat slave Interested in Blockchain - T...	2010-01-12 07:00:04	742.0	716	2444	False	2021-02-10 23:53:30	Trade #Crypto on #Binance \n\n Enjoy #Cashb...	['Crypto', 'Binance', 'Cashback']	Twitter Web App	False

Fig. 3 Sample raw tweets

data), and the data is filled with mean values for these days. The distribution of raw tweets is shown in Fig. 2. Sample raw tweets are also demonstrated in Fig. 3.

Twitter data has no predefined structure, and moreover, it has very noisy texts in general. As a result, to use tweets in sentiment analysis, some extensive preprocessing techniques are applied to the collected Twitter data. First, short tweets (i.e., having less than five words) are ignored and dropped from the data set. Unnecessary fields of data set are also dropped. Tweets are then converted to lower case representations, and hashtags, retweets, non-letters, and emojis are removed. URLs are also removed from original tweets. Tweets are then split into words, and stop words are removed. Finally, stemming is applied to clean tweets. Sample clean tweets are showed in Fig. 4.

3.3 Sentiment Analysis

Sentiments can be simply defined as someone's opinion or feelings about an object, and it is used to determine whether the statement about the object is positive, negative, or neutral. Tweets are often helpful to be used in sentiment analysis research.

In this work, polarity and subjectivity calculation of cleaned tweets are evaluated by using TextBlob. TextBlob is a simple Natural Language Processing (NLP) library that supports complex textual data analysis and operations. Text is represented by

	tweets	cleantext
10000	Indian parliament reportedly considering fast-...	[indian, parliament, reportedli, consid, fast,...
10001	Letsssss gooooooooo #telcoin push 9sat next is...	[letsssss, gooooooooo, telcoin, push, 9sat, ne...
10002	#Bitcoin #BTC current price (GBP): £28,711\nLi...	[bitcoin, btc, current, price, gbp, 28, 711, l...
10003	#Bitcoin W Pattern confirmation!\n\n#BTC #Cryp...	[bitcoin, w, pattern, confirm, btc, crypto, cr...
10004	The #biggest #transaction of all time! \n\nIn...	[biggest, transact, time, 2013, 194, 993, bitc...
10005	Government Digital Transformation - listen her...	[govern, digit, transform, listen, http, co, o...
10006	No Monday Syndrome!\n\nYour #cryptocurrency tr...	[monday, syndrom, cryptocurr, transact, secur,...
10007	Buyer alert: 19 $BTC bought at market @ 395...	[buyer, alert, 19, btc, bought, market, 39500,...
10008	1 BTC Price: Bitstamp 39530.89 USD Coinbase U...	[1, btc, price, bitstamp, 39530, 89, usd, coin...
10009	New Video Alert!! \n\n BITCOIN WEEKLY - 8/2...	[new, video, alert, bitcoin, weekli, 8, 2, 21,...

Fig. 4 Sample clean tweets

bag of words for sentiment analysis. TextBlob assigns individual scores to all these words and then calculates the final sentiment. The polarity and subjectivity of a sentence are returned. The polarity value is defined between [-1,1], -1 for negative and 1 for positive sentiment. After finding the polarity of tweets, some threshold is applied to the polarity values to categorize them as *positive* or *negative*. The example data is illustrated in Fig. 5.

The positive and negative tweets of the same day are grouped together, and the ratios are calculated. The daily percentages are then added as two columns to the tweet data set (i.e., positive and negative ratio columns).

3.4 Proposed Model

This study introduces a novel ensemble model that merges tweet sentiments with historical Bitcoin data to enhance price prediction. The proposed model is a combination of two distinct models. The first model, referred to as EMD-LSTM, incorporates Empirical Mode Decomposition (EMD) into the LSTM network for processing daily Bitcoin price data. The second model, SSLSTM (Stacked Sentiment LSTM), leverages the daily ratio of positive and negative tweets related to Bitcoin, in addition to daily price and volume data. The outputs of both models are subsequently aggregated to generate the final prediction values.

Empirical Mode Decomposition (EMD) is a signal processing method for decomposing original signal into sub-series called intrinsic mode functions (IMFs) along with residue (Huang et al., 1998). EMD is an adaptive unsupervised method and is generally used to handle non-stationary and non-linear data (Bedi & Toshniwal, 2018). The decomposition process produces these IMFs in different time scales and can be applied to any type of signal.

Figure 6 illustrates the sample output of EMD algorithm applied to Bitcoin price data. EMD algorithm is widely used for time series data since it produces IMFs, and the original signal can be reconstructed by using these IMFs without losing any data. Moreover, it can deal with trend patterns in time series data, especially for non-stationary ones. Thus, it can be adapted to Bitcoin price data since it can have various components and is affected by different factors.

	tweets	cleantext	subjectivity	polarity	sentiment
10000	Indian parliament reportedly considering fast-...	[indian, parliament, reportedli, consid, fast,...	0.10	0.0000	neutral
10001	Letsssss gooooooooo #telcoin push 9sat next is...	[letssssss, gooooooooo, telcoin, push, 9sat, ne...	0.00	0.0000	neutral
10002	#Bitcoin #BTC current price (GBP): £28,711\nLi...	[bitcoin, btc, current, price, gbp, 28, 711, l...	0.10	0.0000	neutral
10003	#Bitcoin W Pattern confirmation!\n\n#BTC #Cryp...	[bitcoin, w, pattern, confirm, btc, crypto, cr...	0.10	0.0000	neutral
10004	The #biggest #transaction of all time! \n\nIn...	[biggest, transact, time, 2013, 194, 993, bitc...	0.50	0.4000	positive
10005	Government Digital Transformation - listen her...	[govern, digit, transform, listen, http, co, o...	0.25	-0.0625	negative
10006	No Monday Syndrome!\n\nYour #cryptocurrency tr...	[monday, syndrom, cryptocurr, transact, secur,...	0.00	0.0000	neutral
10007	Buyer alert: 19 $BTC bought at market @ 395...	[buyer, alert, 19, btc, bought, market, 39500,...	0.50	0.6250	positive
10008	1 BTC Price: Bitstamp 39530.89 USD Coinbase U...	[1, btc, price, bitstamp, 39530, 89, usd, coin...	0.50	0.2500	positive
10009	New Video Alert!! \n\n BITCOIN WEEKLY - 8/2...	[new, video, alert, bitcoin, weekli, 8, 2, 21,...	0.00	0.0000	neutral

Fig. 5 Sample tweets with polarity and subjectivity

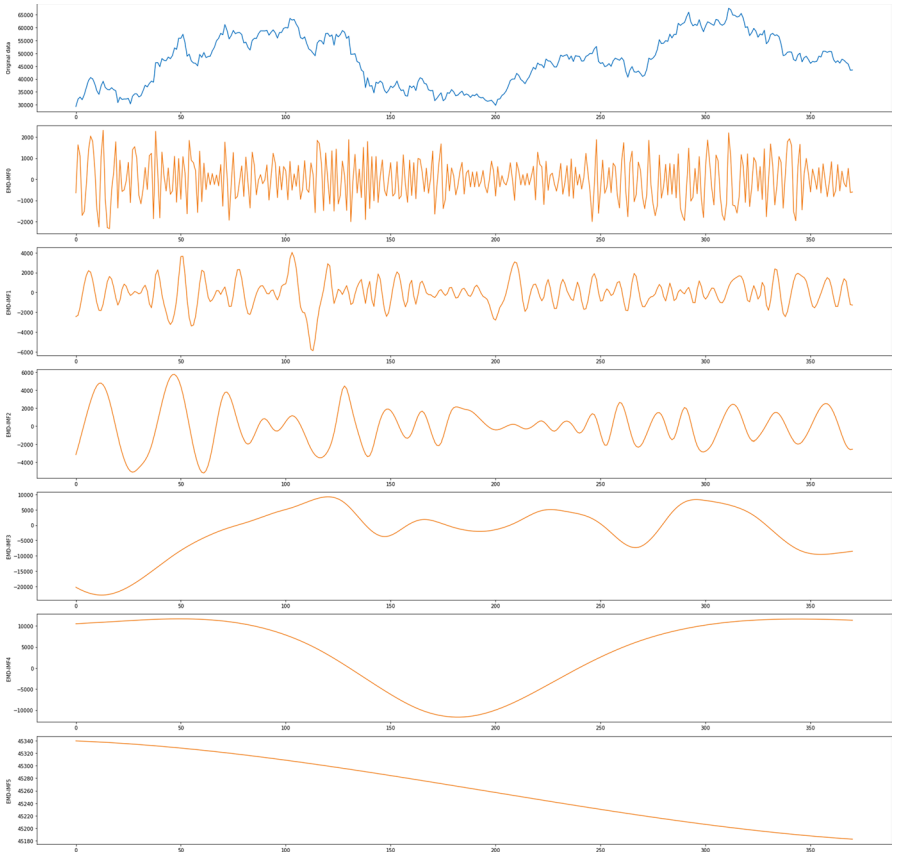

Fig. 6 Sample decomposition of Bitcoin data using EMD

LSTM neural networks are designed as a new recurrent neural network (RNN) form. Essentially, LSTM networks have their memory structure. Typical RNNs endeavor to solve the poor performance of feed-forward neural networks on sequential inputs. They are widely used in speech recognition, opinion and sentiment analysis, text processing, and time series prediction. In the LSTM model, an extended data sequence is memorized or held by adapting a gating structure. The structure of an LSTM cell is depicted in Fig. 7.

The forget gate is utilized to specify which data will be preserved or not. In order to achieve this preservation, the following formula are used;

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

where x_t is input at time t, h_{t-1} is the output of previous cell, and σ is sigmoid function. The information is kept in the cell state if forget gate produces one as output. In the next stage, the sigmoid function constructs a vector. This vector contains possible new values. Input gates are used to specify the updated values, and new possible

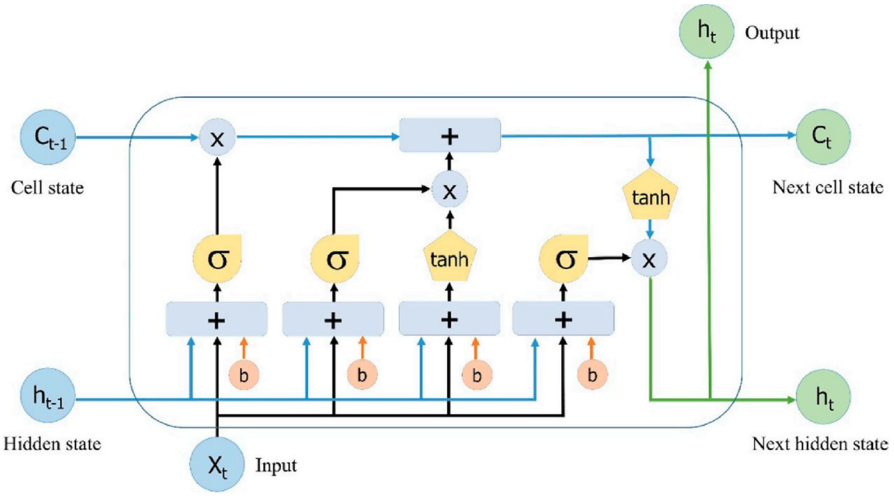

Fig. 7 LSTM cell structure

values are stored in the vector $C_t^{'}$. This new vector is constructed with the following formulas;

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$
$$C_t^{'} = tanh(W_c[h_{t-1}, x_t] + b_c)$$

Now cell's old state C_{t-1} is updated to new cell state C_t.

$$C_t = f_t * C_{t-1} + i_t * C_t^{'}$$

Eventually, we select the network's output regarding on the cell state. This selection process is carried out by using the following formulas;

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$
$$h_t = o_t * tanh(C_t)$$

Stacked LSTMs (SLSTMs) are a special type of classical LSTMs (Graves et al., 2013). In SLSTMs, one and/or more LSTM layers are utilized and merged. The first LSTM layer employs the time series data as input and deliver the output. This output now becomes the input of next LSTM layer. All LSTM layers have an identical inner architecture with various units. Figure 8 represents an example SLSTM network.

In this study, the initial step involves decomposing only the daily Bitcoin price data using the EMD algorithm, resulting in the extraction of six Intrinsic Mode Functions (IMFs). During this phase of the research, tweet sentiments and other Bitcoin-related features (such as average price, opening price, volume, etc.) are not incorporated. Consequently, six separate LSTM models are trained, and their outputs are amalgamated through aggregation. This approach involves

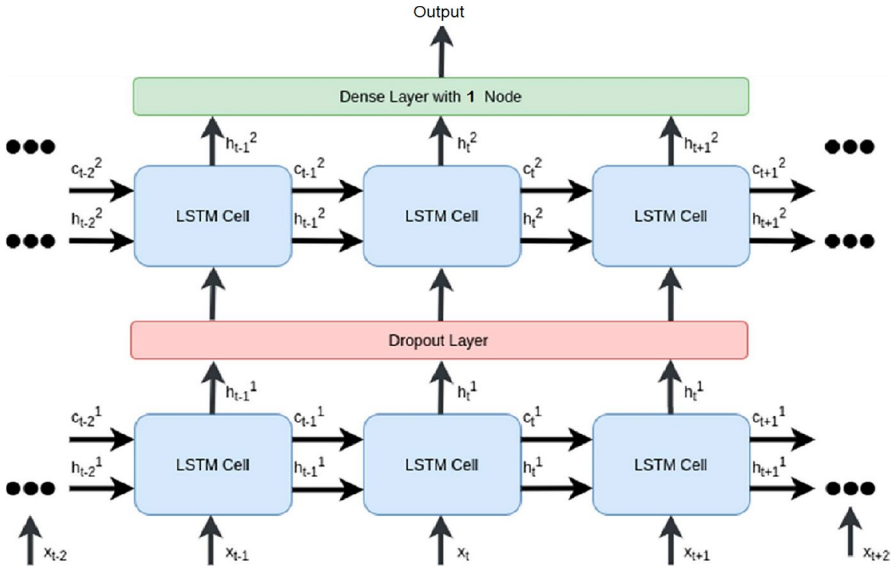

Fig. 8 Stacked LSTM network

subdividing the data into smaller sub-tasks, utilizing LSTMs for prediction, and ultimately combining the outputs to yield optimal prediction results. The steps of this process are described as follows;

- Step 1: The EMD algorithm is applied in order to split time series data into into IMFs and residual parts.
- Step 2: Input of LSTM models are constructed by creating a feature matrix.
- Step 3: This input matrix is split to train and test data.
- Step 4: LSTM models are trained using the train data
- Step 5: Each model output is evaluated to generate final output.

Hence, the author generated Bitcoin price predictions exclusively relying on historical daily price data. Furthermore, a Stacked LSTM model was developed, incorporating both daily price data and sentiment analysis of tweets. The sentiment of tweets was assessed using TextBlob, and the positive and negative tweet ratios were computed for each day. These figures were incorporated into the daily price dataset as two additional columns (*pos* and *neg*). Sample rows are depicted in Fig. 9. Subsequently, Stacked LSTM with sentiment analysis (SSLSTM) was trained on this dataset, yielding prediction results. The outputs from both models (EMD-LSTM and SSLSTM) were then merged by averaging, ultimately producing the final prediction values.

In this study, the author has used Root Mean Square Error (RMSE) and Mean Absolute Error(MAE) metrics to evaluate the proposed model's prediction accuracy. RMSE is calculated by using the following formula;

date	open	high	low	close	Volume BTC	pos	neg
2021-01-01	28923.63	29600.00	28624.57	29331.69	54182.92501	0.45845	0.05001
2021-01-02	29331.70	33300.00	28946.53	32178.33	129993.87340	0.57480	0.05203
2021-01-03	32176.45	34778.11	31962.99	33000.05	120957.56680	0.48389	0.05885
2021-01-04	33000.05	33600.00	28130.00	31988.71	140899.88570	0.38149	0.08008
2021-01-05	31989.75	34360.00	29900.00	33949.53	116049.99700	0.52304	0.08422
2021-01-06	33949.53	36939.21	33288.00	36769.36	127139.20130	0.45125	0.10656
2021-01-07	36769.36	40365.00	36300.00	39432.28	132825.70040	0.47229	0.08922
2021-01-08	39432.48	41950.00	36500.00	40582.81	139789.95750	0.58689	0.10403
2021-01-09	40586.96	41380.00	38720.00	40088.22	75785.97968	0.33868	0.11207
2021-01-10	40088.22	41350.00	35111.11	38150.02	118209.54450	0.43265	0.11524

Fig. 9 Sample input data for sentiment based training

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (Y_i' - Y_i)^2}{n}}$$

where Y_i' is predicted value and Y_i is actual value. The following formula is used for MAE calculation;

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |Y_i' - Y_i|$$

4 Experiments and Results

In the experiments, daily Bitcoin price data from 01-Jan-2021 to 06-Jan-2022 is collected from https://in.investing.com. Bitcoin tweets are crawled using Twitter API and Tweepy. Sentiment analyses are evaluated using TextBlob API. The daily Bitcoin price is illustrated in Fig. 10. It shows that Bitcoin prices have an upward trend. However, the data has been highly volatile and nonlinear in nature.

From Fig. 6, the first two IMFs show high-frequency characteristics, the next two IMFs have medium frequency and the last two show low-frequency characteristics. Thus, these components are grouped together to develop and train corresponding LSTM models (Fig. 11). The train and validation loss of each model is depicted in Figs. 12, 13 and 14. The outputs of each model are also illustrated in Figs. 15, 16 and 17. These outputs are then aggregated to get final prediction values.

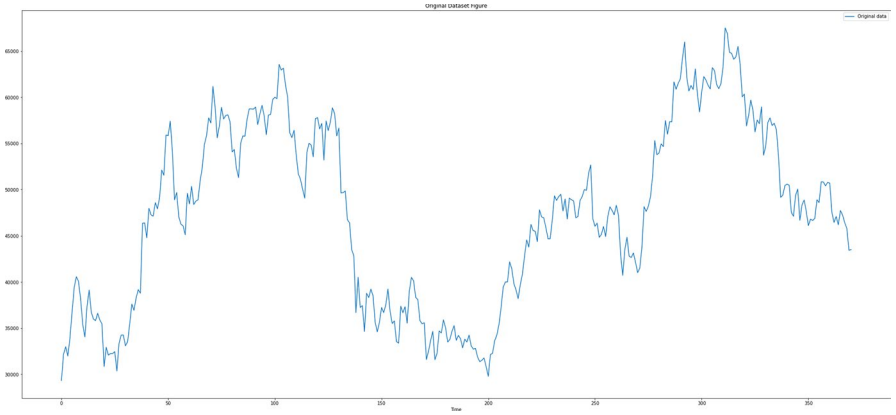

Fig. 10 Daily Bitcoin price data

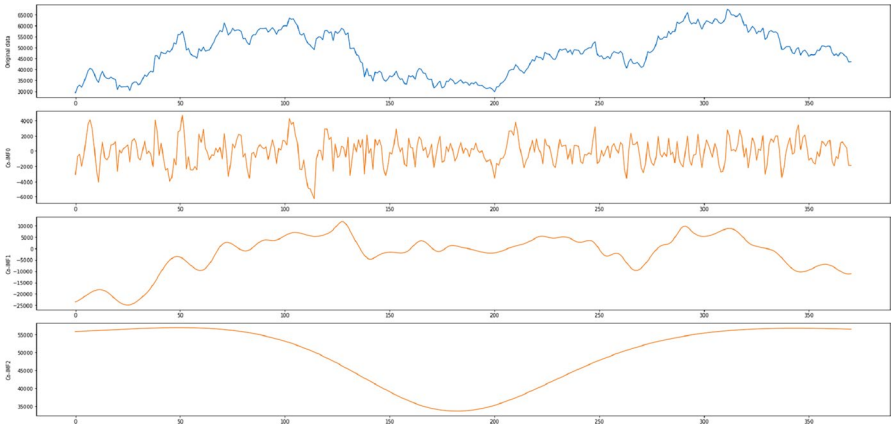

Fig. 11 Groups of EMD components

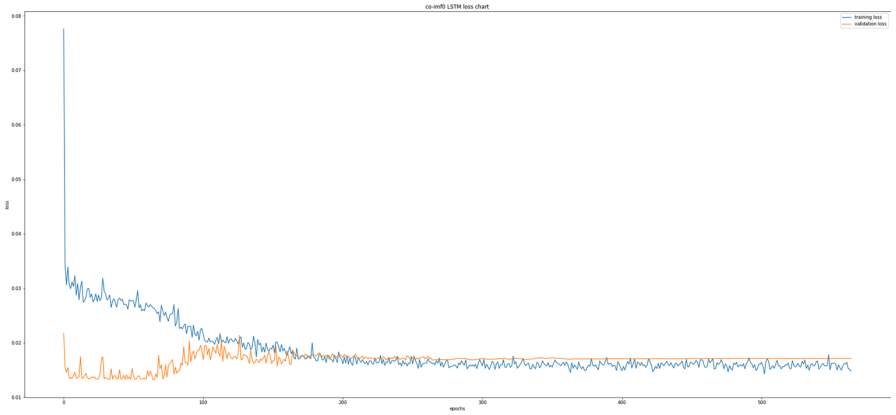

Fig. 12 Train and validation loss graph for high frequency components

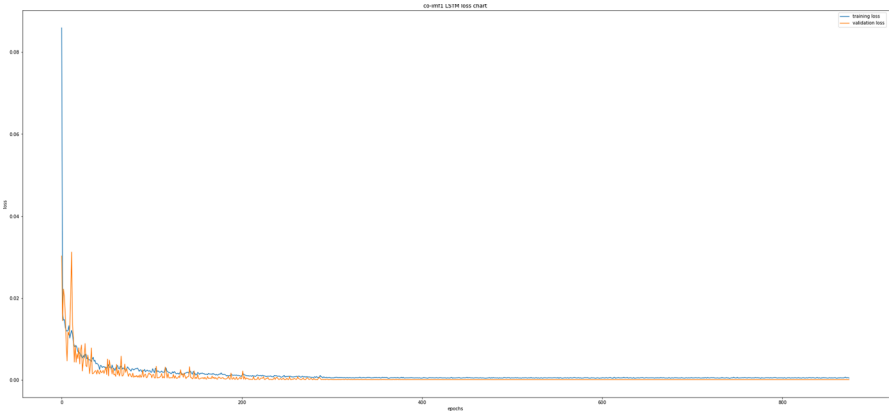

Fig. 13 Train and validation loss graph for medium frequency components

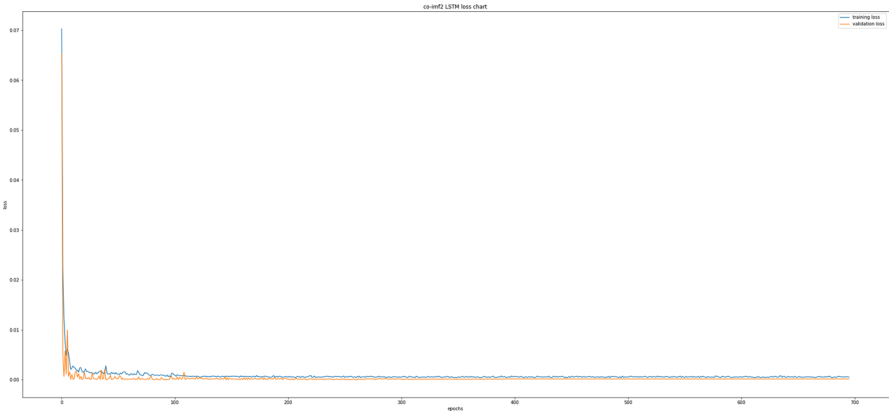

Fig. 14 Train and validation loss graph for low frequency components

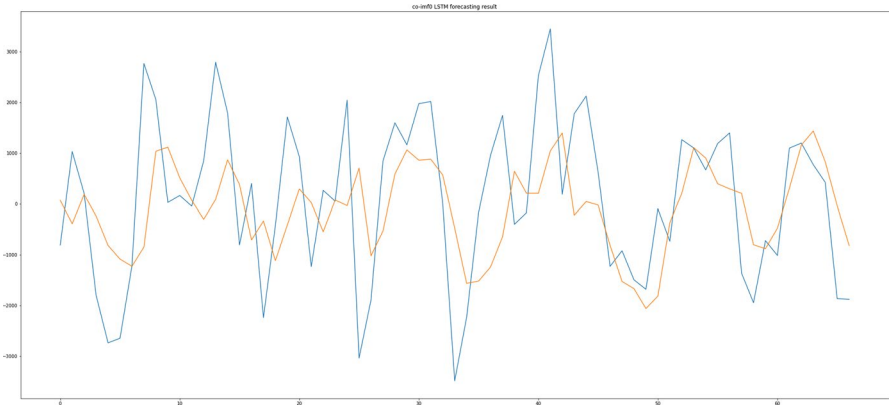

Fig. 15 Forecasting results using high frequency components

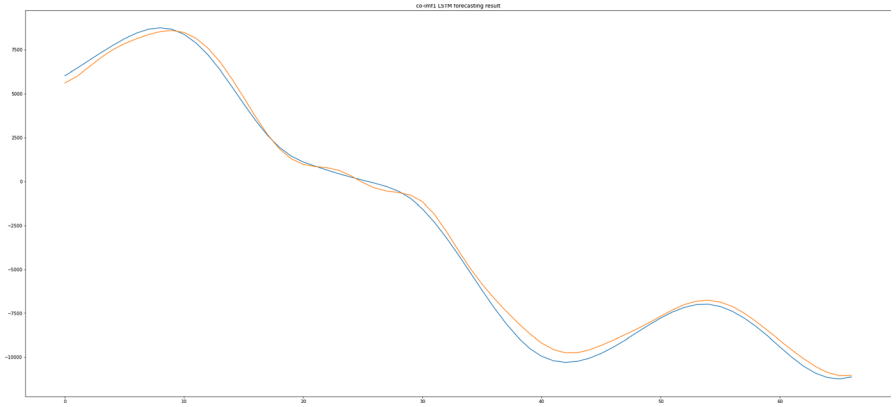

Fig. 16 Forecasting results using medium frequency components

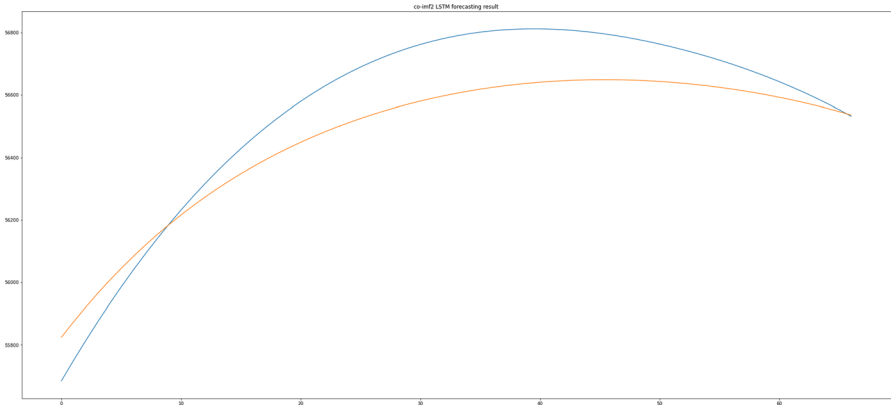

Fig. 17 Forecasting results using low frequency components

In this study, 17 different forecast models were established based on different model structures and input vectors. The details of each model are shown in Table 1. The input vectors of each model type are summarized in Table 2.

To assess the effectiveness of the forecasting models, the dataset is divided into a training set and a testing set with a 4:1 ratio (80% training data and 20% test data). Table 3 provides an overview of the Bitcoin price prediction results. The numbers highlighted in bold indicate that the respective method achieved the best performance for this dataset under the specified performance evaluation index. The test results demonstrate that the proposed method exhibits the lowest MAE and RMSE compared to the other methods. Consequently, incorporating tweet polarity alongside actual price data enhances prediction accuracy. Figure 18 presents the historical Bitcoin price data and predicted values for the next 67 days. Furthermore, Fig. 19 provides a more detailed comparison of actual and predicted values for the test dataset.

Table 1 Model details

Model name	Detail	First layer nodes	Second layer nodes	Batch size	Epochs
$LSTM_{sentimental}$	Sentimental LSTM	256	–	16	100
$GRU_{sentimental}$	Sentimental GRU	256	–	32	100
$SLSTM_{sentimental}$	Sentimental Stacked LSTM	256	128	14	100
$SGRU_{sentimental}$	Sentimental Stacked GRU	256	128	64	100
$BLSTM_{sentimental}$	Sentimental Bidirectional LSTM	256	128	128	100
$BGRU_{sentimental}$	Sentimental Bidirectional GRU	256	128	256	100
$1DCNN\_LSTM_{sentimental}$	Sentimental 1DCNN and LSTM	256	–	64	100
$1DCNN\_GRU_{sentimental}$	Sentimental 1DCNN and GRU	256	–	32	100
$LSTM_{unsentimental}$	Unsentimental LSTM	8	–	7	100
$GRU_{unsentimental}$	Unsentimental GRU	64	–	14	100
$SLSTM_{unsentimental}$	Unsentimental Stacked LSTM	128	64	128	100
$SGRU_{unsentimental}$	Unsentimental Stacked GRU	256	128	14	100
$BLSTM_{unsentimental}$	Unsentimental Bidirectional LSTM	256	128	128	100
$BGRU_{unsentimental}$	Unsentimental Bidirectional GRU	256	128	128	100
$1DCNN\_LSTM_{unsentimental}$	Unsentimental 1DCNN and LSTM	16	–	14	100
$1DCNN\_GRU_{unsentimental}$	Unsentimental 1DCNN and GRU	256	–	14	100
$EMD\_LSTM_{ensemble}$	Proposed model	128	64	30	1000

The test results demonstrated a considerable improvement in Bitcoin price prediction when sentiment analysis using Bitcoin tweets was paired with the EMD structure and deep learning techniques. Three key aspects are responsible for the results improvement: First of all, we utilized behavioral tweet concepts, which take into account the influence of emotional elements on the price of Bitcoin. Secondly, we employed EMD for extracting trend components, resulting in more predictable sequences, and finally, we used deep learning model which focuses on the most crucial information within large datasets.

In summary, the proposed scheme, which combines LSTM with tweet sentiments, and EMD, consistently achieves the highest accuracy, and the closest predictive value when predicting the Bitcoin price. Timely and accurate prediction of Bitcoin price is critical for investor decision-making. More accurate and timely Bitcoin price

Table 2 Model input vectors

Model	Type	Input vector
RNN	Sentimental Models	Opening price, Closing Price, Highest Price, Lowest Price, Volume, pos_{ratio} and neg_{ratio}
RNN	Unsentimental Models	Opening price, Closing Price, Highest Price, Lowest Price, Volume
Proposed Model	Sentimental Stacked LSTM	Opening price, Closing Price, Highest Price, Lowest Price, Volume, pos_{ratio} and neg_{ratio}
Proposed Model	EMD_LSTM	Closing Price

predictions can lead to more timely and reasonable actions and healthy guidance of the cryptocurrency market, contributing to sustainable economic development.

5 Conclusion

The cryptocurrency market has become a popular investment area recently, and social media significantly impacts this market, particularly on Bitcoin. Various models for Bitcoin price prediction have been developed in the past decade. Most recent works used AI-based models to capture the trend and also non-linear patterns. However, these techniques mainly use historical data and cannot capture social media's effect on market trends.

In this study, we present a novel LSTM-based model designed for Bitcoin price prediction. Specifically, our approach incorporates sentiment indices to account for users' emotional tendencies. Additionally, we enhance the LSTM-based model through the use of Empirical Mode Decomposition (EMD), which decomposes the complex Bitcoin pricing sequence into simpler and more predictive sequences. Our model has two different LSTM networks; sentimental LSTM and unsentimental LSTM. The original Bitcoin price data in the former network is decomposed into several components using EMD. These components have formed the input for corresponding inner LSTM models for prediction.

Beyond price data, our model incorporates the sentiments expressed by investors and users, extracted from daily Bitcoin-related tweets through an unsupervised classification technique. We assess the polarity ratio and combine it with daily price and volume data to construct a sentiment-based LSTM model. Subsequently, we merge these two LSTM models to generate our final predictions.

We conducted a comprehensive comparison of the proposed model against 16 different models, wherein eight of them solely utilize price and volume data, while the remaining eight models incorporate polarity information derived from historical data. All of these models are based on RNN networks, encompassing LSTMs and GRUs. Our test results unequivocally affirm that the proposed ensemble model outperforms all of the RNN-based models. Notably, our model demonstrates the capability to effectively capture the impact of social media trends on Bitcoin price prediction. Furthermore, our model extends its support for

Table 3 Comparisons of predicted results

Model	MAE	RMSE
$LSTM_{sentimental}$	1572.640638	1985.763516
$GRU_{sentimental}$	1552.923081	1990.844399
$SLSTM_{sentimental}$	1683.756338	2139.514042
$SGRU_{sentimental}$	1678.488343	2028.370489
$BLSTM_{sentimental}$	2305.463599	2815.457879
$BGRU_{sentimental}$	1874.13637	2381.182112
$1DCNN\_LSTM_{sentimental}$	2211.135163	2711.501403
$1DCNN\_GRU_{sentimental}$	2052.351429	2590.692901
$LSTM_{unsentimental}$	1577.775951	2081.693218
$GRU_{unsentimental}$	1407.409122	1893.589101
$SLSTM_{unsentimental}$	1685.997039	2242.583125
$SGRU_{unsentimental}$	1560.115877	1900.501783
$BLSTM_{unsentimental}$	1918.141045	2425.092162
$BGRU_{unsentimental}$	1583.108612	2139.431157
$1DCNN\_LSTM_{unsentimental}$	2394.498398	2942.648354
$1DCNN\_GRU_{unsentimental}$	2278.329132	2742.75101
$EMD\_LSTM_{ensemble}$	990.6232523	1282.33121

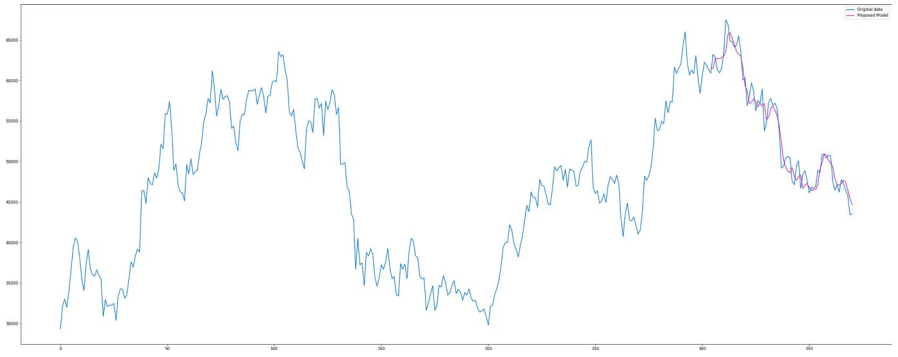

Fig. 18 Bitcoin price data and predicted values

sentiment classification across various social media platforms, including Twitter, Facebook, and others, as it leverages the daily ratio of positive and negative opinions.

As a result, the proposed scheme outlined in this paper demonstrates significant potential for contributing to the benefit of the countries by offering guidance to their governments in rationalizing and regulating the cryptocurrency market. It also has the potential to assist individuals in making profitable investment decisions.

Our future endeavors will involve the application of additional supervised methods to further enhance prediction accuracy. Furthermore, we intend to integrate various data sources for collecting and analyzing user comments and opinions. It

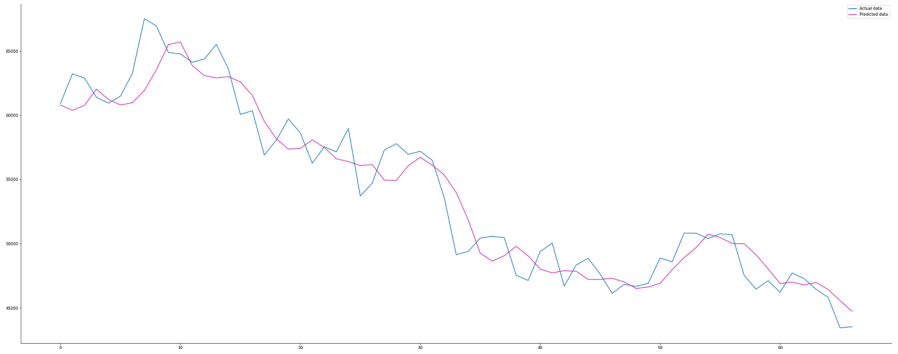

Fig. 19 Actual and predicted values for test data set

is reasonable to anticipate that the combination of supervised models and diverse data sources for training the networks will enable us to achieve even superior performance.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). The research received no external funding.

Data availability The datasets used and/or analyzed during the current study and codes are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no competing interests

References

Aggarwal, D., Chandrasekaran, S., & Annamalai, B. (2020). A complete empirical ensemble mode decomposition and support vector machine-based approach to predict Bitcoin prices. *Journal of Behavioral and Experimental Finance, 27*, 100335. https://doi.org/10.1016/j.jbef.2020.100335

Bedi, J., & Toshniwal, D. (2018). Empirical mode decomposition based deep learning for electricity demand forecasting. *IEEE Access, 6*, 49144–49156. https://doi.org/10.1109/ACCESS.2018.2867681

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

Chen, L., Chi, Y., Guan, Y., & Fan, J. (2019). A hybrid attention-based EMD-LSTM model for financial time series prediction. In *2019 2nd international conference on artificial intelligence and big data, ICAIBD 2019* (pp. 113–118). https://doi.org/10.1109/ICAIBD.2019.8837038.

Chen, W., Xu, H., Jia, L., & Gao, Y. (2021). Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants. *International Journal of Forecasting, 37*(1), 28–43. https://doi.org/10.1016/j.ijforecast.2020.02.008

Cocco, L., Tonelli, R., & Marchesi, M. (2021). Predictions of bitcoin prices through machine learning based frameworks. *PeerJ Computer Science, 7*, 1–23. https://doi.org/10.7717/PEERJ-CS.413

Derbentsev, V., Babenko, V., Khrustalev, K., Obruch, H., & Khrustalova, S. (2021). Comparative performance of machine learning ensemble algorithms for forecasting cryptocurrency prices. *International Journal of Engineering, Transactions A: Basics, 34*(1), 140–148. https://doi.org/10.5829/IJE.2021.34.01A.16

Dokur, E., Kurban, M., & Ceyhan, S. (2016). Hybrid model for short term wind speed forecasting using empirical mode decomposition and artificial neural network. In *ELECO 2015—9th international conference on electrical and electronics engineering (ii)* (pp. 420–423). https://doi.org/10.1109/ELECO.2015.7394591.

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research, 270*(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054

Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of Twitter sentiment analysis methods. *ACM Computing Surveys*. https://doi.org/10.1145/2938640

Graves, A., Mohamed, A.-r., & Hinton, G. E. (2013). Speech Recognition with Deep Recurrent Neural Networks. CoRR **abs/1303.5**. arXiv:1303.5778.

Gurrib, I., & Kamalov, F. (2021). Predicting bitcoin price movements using sentiment analysis: a machine learning approach. Studies in Economics and Finance **ahead-of-p**(ahead-of-print). https://doi.org/10.1108/SEF-07-2021-0293.

Gurrib, I., & Kamalov, F. (2022). Predicting bitcoin price movements using sentiment analysis: A machine learning approach. *Studies in Economics and Finance, 39*(3), 347–364. https://doi.org/10.1108/SEF-07-2021-0293

Gyamerah, S. A. (2019). Are Bitcoins price predictable? Evidence from machine learning techniques using technical indicators. arXiv:1909.01268.

Hamayel, M. J., & Owda, A. Y. (2021). A novel cryptocurrency price prediction model using GRU, LSTM and bi-LSTM machine learning algorithms. *AI, 2*(4), 477–496. https://doi.org/10.3390/ai2040030

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Snin, H. H., Zheng, Q., Yen, N. C., Tung, C. C., & Liu, H. H. (1998). The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 454*(1971), 903–995. https://doi.org/10.1098/rspa.1998.0193

Ji, S., Kim, J., & Im, H. (2019). A comparative study of bitcoin price prediction using deep learning. *Mathematics*. https://doi.org/10.3390/math7100898

Jin, X., Zhu, K., Yang, X., & Wang, S. (2021). Estimating the reaction of Bitcoin prices to the uncertainty of fiat currency. *Research in International Business and Finance, 58*, 101451. https://doi.org/10.1016/j.ribaf.2021.101451

Jin, Z., Yang, Y., & Liu, Y. (2020). Stock closing price prediction based on sentiment analysis and LSTM. *Neural Computing and Applications, 32*(13), 9713–9729. https://doi.org/10.1007/s00521-019-04504-2

Keceli, A. S., Catal, C., Kaya, A., & Tekinerdogan, B. (2020). Development of a recurrent neural networks-based calving prediction model using activity and behavioral data. *Computers and Electronics in Agriculture, 170*, 105285. https://doi.org/10.1016/j.compag.2020.105285

Khedr, A. M., Arif, I., Pravija Raj, P. V., El-Bannany, M., Alhashmi, S. M., & Sreedharan, M. (2021). Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey. *Intelligent Systems in Accounting, Finance and Management, 28*(1), 3–34. https://doi.org/10.1002/isaf.1488

Ko, C. R., & Chang, H. T. (2021). LSTM-based sentiment analysis for stock price forecast. *PeerJ Computer Science, 7*, 1–23. https://doi.org/10.7717/peerj-cs.408

Kraaijeveld, O., & De Smedt, J. (2020). The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. *Journal of International Financial Markets, Institutions and Money, 65*, 101188. https://doi.org/10.1016/j.intfin.2020.101188

Kristoufek, L. (2015). What are the main drivers of the bitcoin price? Evidence from wavelet coherence analysis. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0123923. arXiv:1406.0268.

Lahmiri, S., Saade, R. G., Morin, D., & Nebebe, F. (2020). An artificial neural networks based ensemble system to forecast bitcoin daily trading volume. In *Proceedings of 2020 5th international conference on cloud computing and artificial intelligence: Technologies and applications, CloudTech 2020. Institute of Electrical and Electronics Engineers Inc.*https://doi.org/10.1109/CloudTech49835.2020.9365913.

Lara-Benítez, P., Carranza-García, M., & Riquelme, J. C. (2021). An experimental review on deep learning architectures for time series forecasting. *International Journal of Neural Systems*. https://doi.org/10.1142/S0129065721300011. arXiv:2103.12057.

Lim, B., & Zohren, S. (2021). Time-series forecasting with deep learning: A survey. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. https://doi.org/10.1098/rsta.2020.0209. arXiv:2004.13408.

Livieris, I. E., Pintelas, E., Stavroyiannis, S., & Pintelas, P. (2020). Ensemble Deep learning models for forecasting cryptocurrency time-series. *Algorithms, 13*(5), 1–21. https://doi.org/10.3390/A13050121

Mohapatra, S., Ahmed, N., & Alencar, P. (2019). KryptoOracle: A real-time cryptocurrency price prediction platform using twitter sentiments. In *Proceedings—2019 IEEE international conference on big data, big data 2019* (pp. 5544–5551). https://doi.org/10.1109/BigData47090.2019.9006554. arXiv:2003.04967.

Mtiraoui, A., Boubaker, H., & BelKacem, L. (2023). A hybrid approach for forecasting bitcoin series. *Research in International Business and Finance, 66*, 102011. https://doi.org/10.1016/j.ribaf.2023.102011

Munim, Z. H., Shakil, M. H., & Alon, I. (2019). Next-day bitcoin price forecast. *Journal of Risk and Financial Management, 12*(2), 103. https://doi.org/10.3390/jrfm12020103

Pintelas, E., Livieris, I., Stavroyiannis, S., Kotsilieris, T., & Pintelas, P. (2020). Fundamental research questions and proposals on predicting cryptocurrency prices using DNNs (February), pp. 1–20.

Prajapati, P. (2020). Predictive analysis of Bitcoin price considering social sentiments. arXiv:2001.10343.

Roy, S., Nanjiba, S., & Chakrabarty, A. (2018). Bitcoin price forecasting using time series analysis. In *2018 21st international conference of computer and information technology (ICCIT)* (pp. 1–5). https://doi.org/10.1109/ICCITECHN.2018.8631923.

Shin, M. J., Mohaisen, D., & Kim, J. (2021). Bitcoin Price Forecasting via Ensemble-based LSTM Deep Learning Networks. In *International conference on information networking, vol. 2021-Janua* (pp. 603–608). IEEE Computer Society. https://doi.org/10.1109/ICOIN50884.2021.9333853.

Zhang, D., Sun, Y., Duan, H., Hong, Y., & Wang, S. (2023). Speculation or currency? Multi-scale analysis of cryptocurrencies-The case of Bitcoin. *International Review of Financial Analysis, 88*, 102700. https://doi.org/10.1016/j.irfa.2023.102700

Zhao, D., Rinaldo, A., & Brookins, C. (2019). Cryptocurrency price prediction and trading strategies using support vector machines (January 2009). arXiv:1911.11819.

Zhou, F., Huang, Z., & Zhang, C. (2022). Carbon price forecasting based on CEEMDAN and LSTM. *Applied Energy, 311*, 118601. https://doi.org/10.1016/j.apenergy.2022.118601

Zimbra, D., Abbasi, A., Zeng, D., & Chen, H. (2018). The state-of-the-art in twitter sentiment analysis: A review and benchmark evaluation. *ACM Transactions on Management Information Systems*. https://doi.org/10.1145/3185045